1 Hierarchical ℓ_1/ℓ_q-norms

$$\min_{\boldsymbol{w}\in\mathbb{R}^p} \quad \frac{1}{2}\|\boldsymbol{u}-\boldsymbol{w}\| + \mu\Omega(\boldsymbol{w}) \tag{1.1}$$

$$\max_{\boldsymbol{v}\in\mathbb{R}^p} \quad -\frac{1}{2}\left[\|\boldsymbol{u}-\boldsymbol{v}\|_2^2 - \|\boldsymbol{u}\|^2\right] \quad \text{such that} \quad \Omega^*(\boldsymbol{v}) \le \mu \tag{1.2}$$

Proximal operator of Hierarchical norms

次のノルムを考え：$\Omega : \boldsymbol{w} \to \sum_{g\in\mathcal{G}} \|\boldsymbol{w}_g\|_q$, with $q \in \{2, \infty\}$，$\mathcal{G}$ は木構造のグループ集合である．$\preceq$ 記号を次のように定義する．

$$g_1 \preceq g_2 \Rightarrow \{g_1 \subseteq g_2 \ \text{ or } \ g_1 \cap g_2 = \emptyset\}$$

$g_1 \preceq \cdots \preceq g_m$，$m = |\mathcal{G}|$, $\underset{g}{\mathrm{Prox}} := \boldsymbol{w}_g \mapsto \underset{\mu\|\cdot\|_q}{\mathrm{Prox}}(\boldsymbol{w}_g)$ とおくと，Hierarchical ℓ_1/ℓ_q-norms の proximal operator は

$$\underset{\mu\Omega}{\mathrm{Prox}} = \underset{g_m}{\mathrm{Prox}} \circ \cdots \circ \underset{g_1}{\mathrm{Prox}}$$

補題 1.1. ℓ_1, ℓ_q-ノルムの双対ノルムは $\ell_\infty.\ell_{q^*}$-ノルム，$\frac{1}{q} + \frac{1}{q^*} = 1$

補題 1.2. 双対問題と最適解条件

双対問題

$$\max_{\boldsymbol{\xi}\in\mathbb{R}^{p\times|\mathcal{G}|}} -\frac{1}{2}\Big(\|\boldsymbol{u} - \sum_{g\in\mathcal{G}} \boldsymbol{\xi}^g\|_2^2 - \|\boldsymbol{u}\|_2^2\Big) \ \ s.t.\ \forall g\in\mathcal{G}, \|\boldsymbol{\xi}^g\|_* \le \mu \ and \ \boldsymbol{\xi}_j^g = 0 \ if\ j \ne g \tag{1.3}$$

最適解条件

$$\begin{cases} \boldsymbol{w} = \boldsymbol{u} - \sum_{g\in\mathcal{G}} \boldsymbol{\xi}^g \\ \forall g \in \mathcal{G}, \ \ \boldsymbol{\xi}^g = \Pi_\mu^*(\boldsymbol{w}_g + \boldsymbol{\xi}^g) \end{cases} \tag{1.4}$$

Algorithm 1 Block coordinate descent algorithm

input: $\boldsymbol{u} \in \mathbb{R}^p$ とグループ集合 $\mathcal{G}$

Outputs: $(\boldsymbol{w}, \boldsymbol{\xi})$(主双対最適解)

Initialization: $\boldsymbol{w} = \boldsymbol{u}, \boldsymbol{\xi} = 0$

WHILE (maximum number of iterations not reached) DO

  FOR $g \in \mathcal{G}$ DO

    $\boldsymbol{w} \leftarrow \boldsymbol{u} - \sum_{h\ne g} \boldsymbol{\xi}^h$

    $\boldsymbol{\xi}^g \leftarrow \Pi_\mu^*(\boldsymbol{w}_g)$

  END FOR

END WHILE

$\boldsymbol{w} \leftarrow \boldsymbol{u} - \sum_{g\in\mathcal{G}} \boldsymbol{\xi}^g$

補題 1.3. *Alogorithm 1* で，任意の $g \in \mathcal{G}, h \in \mathcal{G}, g \preceq h$ に対し，もし $\boldsymbol{\xi}^g$ を $\boldsymbol{\xi}^h$ の前に更新すれば，$\boldsymbol{\xi}^h$ は $\boldsymbol{\xi}^g$ の最適解条件に影響を与えない．

Algorithm 2 Practical Computation of the Proximal Operator for ℓ_2- or ℓ_∞-norms.

input: $\boldsymbol{u} \in \mathbb{R}^p$ と木構造グループ集合 $\mathcal{G}$

Outputs: $(\boldsymbol{w})$(主問題最適解)

Initialization: $\boldsymbol{w} = \boldsymbol{u}$

FOR $g \in \mathcal{G}$,following the order $\preceq$ DO

$\boldsymbol{w}_g \leftarrow \boldsymbol{w}_g - \Pi_{\|\cdot\|_* \leq \mu}(\boldsymbol{w}_g) = \underset{\mu\|\cdot\|_q}{\mathrm{Prox}}(\boldsymbol{w}_g)$

END FOR

2 Combined $\ell_1 + \ell_1/\ell_q$-norm(sparse group Lasso)

group Lasso

データサンプルは $\boldsymbol{x}_j \in \mathbb{R}^m, j = 1, \ldots, n$，辞書 (Dictionary) は $\boldsymbol{D} \in \mathbb{R}^{m \times p}, \boldsymbol{D} = [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_p]$，$\boldsymbol{x}_j$ は $\boldsymbol{x}_j = \boldsymbol{D}\boldsymbol{\alpha}_j + \epsilon$ と表せる，$\alpha \in \mathbb{R}^p$ は辞書の線形結合重み，$\epsilon \in \mathbb{R}^m$ はノイズである．この問題の group Lasso 解法は

$$\min_{\boldsymbol{a}} \frac{1}{2}\|\boldsymbol{x}_j - \boldsymbol{D}\boldsymbol{\alpha}\|_2^2 + \mu\Omega_{\mathcal{G}}(\boldsymbol{\alpha}) \tag{2.1}$$

この問題に対するの Hierarchical Lasso は

$$\min_{\boldsymbol{a}} \frac{1}{2}\|\boldsymbol{x}_j - \boldsymbol{D}\boldsymbol{\alpha}\|_2^2 + \mu_1\Omega_{\mathcal{G}}(\boldsymbol{\alpha}) + \mu_2\|\boldsymbol{\alpha}\|_1 \tag{2.2}$$

3 Overlapping ℓ_1/ℓ_∞-norms

overlap があるが木構造のないグラーフ集合の Proximal operator を算出するのは難しいだが，$q = \infty$ だけは効率的に計算できる．

参考文献

[1] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach, ”Proximal methods for sparse hierarchical dictionary learning,” Proc. ICML, 2010.

[2] Proximal Methods for Sparse Hierarchical Dictionary Learning: Supplementary Materials `http://www.di.ens.fr/~jenatton/paper/SupplementaryMaterialsICML2010.pdf`

[3] S. Sra, Fast Projections onto $\ell_{1,q}$-Norm Balls for Grouped Feature Selection Lecture Notes in Computer Science, 2011, Volume 6913, Machine Learning and Knowledge Discovery in Databases, Pages 305-317